APPENDIX A
MODELING CONSUMER PREFERENCES FOR GREEN BAY FISHING DAYS AND FISHING DAYS USING STATED AND REVEALED PREFERENCE DATA

A.1 INTRODUCTION

The purpose of this research is to estimate the parameters in two conditional indirect utility functions: one for a Green Bay fishing day, and one for fishing elsewhere. There are two types of data available: stated preference (SP) data and revealed preference (RP) data. The RP data consist of the total number of fishing days for each individual in the sample and the number of those days to Green Bay under current conditions. The SP data consist of the answers to choice questions. Each sampled individual indicated his or her choice between a pair of Green Bay alternatives (Green Bay under different conditions), and then indicated the number of times in n choice occasions (fishing days) the preferred Green Bay alternative would be chosen, in a choice set that includes it and all non-Green Bay fishing sites. For each sampled individual, these two questions are repeated J times, where the characteristics of the Green Bay alternatives in the pairs are varied over the J pairs.

Section A.2 develops the choice probabilities for the two Green Bay alternatives using only the part of the SP data that indicate which Green Bay alternative is chosen. Section A.3 uses all of the SP data and the RP data on the total number of fishing days under current conditions to model how often the preferred Green Bay alternative would be chosen versus some other non-Green Bay site. Section A.4 incorporates the RP data on the total number of fishing days to Green Bay under current conditions, and Section A.5 presents the likelihood function for the model. Section A.6 provides details on the derivation of the probability of choosing the preferred Green Bay alternative over fishing elsewhere, conditional on the utility from the preferred Green Bay alternative being greater than the utility from the Green Bay alternative not chosen.

A.2 CHOICE PROBABILITIES FOR SP GREEN BAY PAIRS

Let utility for the Green Bay alternatives be given by:

$$U_{ij}^{k_{ij}} = b_i' x_{ij}^{k_{ij}} + e_{ij}^{k_{ij}}, \; i = 1, \ldots, m; \; j = 1, \ldots, J; \; k_{ij} \in [1, 2], \tag{1}$$

where $U_{ij}^{k_{ij}}$ is the utility of the k-th alternative of pair j to individual i. That is, i indexes the m respondents, j indexes the J pairs, and k_{ij} indicates which of the two alternatives within each pair is chosen. The $L \times 1$ vector $x_{ij}^{k_{ij}}$ contains the characteristics of the alternatives, and hence the elements of the unknown $L \times 1$ vector b can be interpreted as marginal utilities. The first element of $x_{ij}^{k_{ij}}$ is the difference between choice-occasion income for individual i and the cost of alternative k_{ij}, and the model is restricted to one with a constant marginal utility of money, which is the first element of b.[1] This specification implies no income effects; that is, the probability of choosing any alternative is independent of income. The term $b_i' x_{ij}^{k_{ij}}$ is the nonstochastic part of utility, while $e_{ij}^{k_{ij}}$ represents a stochastic component. The following assumptions are made:

Assumption 1. $b_i = b$ for all i (heterogeneity in the marginal utilities will be considered later); and

Assumption 2. $e_{ij}^{k_{ij}}$ are independent (across i) and identically distributed mean zero normal random variables, uncorrelated with $x_{ij}^{k_{ij}}$, with constant unknown variance s_e^2.

For SP data, it is assumed that the individual does not know his stochastic component before actually deciding on the particular alternative. That is, $e_{ij}^{k_{ij}}$ is assumed to be the sum of factors unknown to *both* the individual and the investigator.[2] Let $K_{ij} \in [1,2]$ be the Bernoulli random variable that is the choice for individual i on occasion j. The individual is assumed to choose alternative k_{ij} with the probability:[3]

$$P(K_{ij} = k_{ij}) = P_{ij}^{k_{ij}} = P(U_{ij}^{k_{ij}} > U_{ij}^{3-k_{ij}}), \tag{2}$$

where k_{ij} is the observed value of K_{ij}. That is, we may think of the individual's choice as a drawing from a Bernoulli distribution with the probability given by Equation 2.

1. Later, different types of individuals are allowed to have different marginal utilities of money.

2. For RP data, the usual discrete-choice model specification is that the disturbances are known to the individual, and the behavioral assumption is utility maximization. This assumption is also sometimes made for SP data, although the rationale is less clear. Burtless and Hausman (1978) and Moffitt (1986) interpret disturbances unknown to the decision-maker in models with piecewise-linear budget constraints. In those models (and the water demand model of Hewitt and Hanneman, 1995) there is also "heterogeneity" error, which is observed by the decision-maker but unobservable to the investigator. Under the assumption that disturbances are known to the individual a priori, they would perform the conceptual experiment of generating n_i pairs of disturbances and evaluating utility for the two scenarios under the assumption of utility maximization. However, this would produce the identical likelihood.

3. In this notation, if the individual chooses alternative $K_{ij} = 1$ [or 2], then the alternative that was not chosen is $3 - K_{ij} = 2$ [or 1].

From Equations 1 and 2 and assumption 1, the probability of choosing alternative k_{ij} is:

$$\begin{aligned} P_{ij}^{k_{ij}} &= P(b' x_{ij}^{k_{ij}} + e_{ij}^{k_{ij}} > b' x_{ij}^{3-k_{ij}} + e_{ij}^{3-k_{ij}}) \\ &= P\left[e_{ij}^{3-k_{ij}} - e_{ij}^{k_{ij}} < -b'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}})\right] \\ &= \Phi\left[-b'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}}) / \sqrt{2}s_e\right] \end{aligned} \tag{3}$$

where $\sqrt{2}s_e$ is the standard deviation of $e_{ij}^{3-k_{ij}} - e_{ij}^{k_{ij}}$ under assumption 2 and $\Phi(\cdot)$ is the univariate standard normal cumulative distribution function. Note that Equation 3 would be the probability in the usual probit model for dichotomous choice under the assumption the individual knows the random component and maximizes utility. This probability will enter into the likelihood function in Section A.5. The parameter vector b is identified only up to the scale factor $\sqrt{2}s_e$, and s_e is not identified, since only the sign and not the scale of the dependent variable (the utility difference) is observed. Nevertheless, we have chosen to list the parameters of the likelihood function (b, s_e) separately. Notice also the *J* observations for each respondent have simply been stacked to produce a data set with *Jm* observations.

A.3 FREQUENCY OF SELECTING THE PREFERRED GREEN BAY ALTERNATIVE VERSUS ANOTHER SITE

Now suppose in addition to the data on k_{ij}, the individual answers a question giving the number of times Green Bay alternative k_{ij} would be chosen compared to some other (non-Green Bay) alternative, in their next n_i choice occasions (fishing days). Utility for the "other" alternative, U_{ij}^0 (fishing elsewhere), is given by Equation 4:

$$U_{ij}^0 = b' x_{ij}^0 + e_{ij}^0, \tag{4}$$

where e_{ij}^0 are disturbances and x_{ij}^0 are the characteristics of the other site. The following assumption characterizes the disturbances:

Assumption 3: The e_{ij}^0 are independent (across *i*) and identically distributed normal random variables, with zero expectation, and $E\left(e_{ij}^0 e_{ij}^{k_{ij}}\right) = s_{e0}$.

In this model, the value of a random variable N_{ij} is known, where N_{ij} is the number of times Green Bay site k_{ij} is chosen over the non-Green Bay site in the next n_i occasions.[4] The nonstochastic parts of the utilities for the two alternatives in this choice set are $b' x_{ij}^{k_{ij}}$ and $b' x_{ij}^0$. The individual knows these, but does not know the random component associated with either alternative because he must decide in advance how he feels at the time of the choice. If $b' x_{ij}^0 < b' x_{ij}^{k_{ij}}$, for example,

4. The parameter n_i is set equal to the number of days individual *i* fished in 1998.

he knows, on average, he would be better off choosing Green Bay site k_{ij} over fishing elsewhere, but he cannot be certain. For some trips, e_{ij}^{0} may be sufficiently larger than $e_{ij}^{k_{ij}}$ so that $U_{ij}^{0} > U_{ij}^{k_{ij}}$. Over a future set of choice occasions, then, it is assumed that he calculates his answer to the number question probabilistically. That is, he calculates the conditional probability that he will prefer alternative k_{ij} over fishing elsewhere (see Equation 6 below) and then reports the closest integer to n_i times that probability.[5] This is the expected number of trips under the assumption made below that the N_{ij} are distributed binomially, and this average number of trips is elicited in the survey.

Since the N_{ij} are counts ranging from zero to n_i, given the behavioral assumption discussed above a plausible stochastic model for the N_{ij} is that they are distributed binomially, $N_{ij} \sim B(n_i, p_{ij}^{0})$, with probability mass function (conditional on the choice of k_{ij}):

$$P(N_{ij} = n_{ij} | K_{ij} = k_{ij}) = \binom{n_i}{n_{ij}} (p_{ij}^{0})^{n_{ij}} (1 - p_{ij}^{0})^{n_i - n_{ij}}, \tag{5}$$

where n_{ij} are the observed values of N_{ij}.[6] Equation 5 will enter into the likelihood function in Section A.5.

The parameter p_{ij}^{0} in Equation 5 is the probability of choosing Green Bay alternative k_{ij} over the "other" site, conditional on choosing alternative k_{ij} over alternative 3 - k_{ij}:

$$\begin{aligned} p_{ij}^{0} &= P(U_{ij}^{k_{ij}} > U_{ij}^{0} | U_{ij}^{k_{ij}} > U_{ij}^{3-k_{ij}}) \\ &= P[e_{ij}^{0} - e_{ij}^{k_{ij}} < -b'(x_{ij}^{0} - x_{ij}^{k_{ij}}) | e_{ij}^{3-k_{ij}} - e_{ij}^{k_{ij}} < -b'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}})] \\ &= \frac{\Phi_2[-b'(x_{ij}^{0} - x_{ij}^{k_{ij}}) / s_{0-e},\ -b'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}}) / \sqrt{2}\, s_e ; r]}{\Phi[-b'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}}) / \sqrt{2}\, s_e]} \end{aligned} \tag{6}$$

where $s_{0-e}^{2} = \mathrm{Var}\,(e_{ij}^{0} - e_{ij}^{k_{ij}}) = s_0^{2} + s_e^{2} - 2s_{e0}$

5. This is in contrast to assuming the individual repeatedly applies a maximum expected utility decision rule, which would imply a corner solution at either zero or n_i. When we consider revealed preference choices below, we assume that the individual knows his stochastic component and maximizes utility.

6. We are effectively assuming: $P(N_{i1} = n_{i1}, \ldots, N_{iJ}) = \prod_{j=1}^{J} P(N_{ij} = n_{ij})$. A model that explicitly recognizes the fact that the same individual makes all n_{ij} choices exists (it is called the multivariate binomial distribution; see Johnson et al., 1997). It appears to be unwieldy, except possibly for the case $J = 2$. For most if not all formulations of this multivariate distribution the marginals are univariate binomial, so the method of estimating the p_{ij}^{0}'s suggested here is justified. There has been some interest in testing the equality of the p_{ij}^{0} across j (see Westfall and Young, 1989), but that is not the main focus here.

and where r is the correlation between $e_{ij}^{0} - e_{ij}^{k_{ij}}$ and $e_{ij}^{3-k_{ij}} - e_{ij}^{k_{ij}}$,

$$r = \frac{s_e^2}{\sqrt{2 s_e^2 s_{0-e}^2}} \tag{7}$$

and Φ and Φ_2 are the standard univariate and bivariate normal distribution functions, respectively.[7] (For details of the derivation of Equation 6, see Section A.6.)

A.4 INCORPORATING THE RP DATA ON ACTUAL GREEN BAY FISHING DAYS

In addition to the SP data and the n_i, we have for each i the number of fishing days to Green Bay, n_i^G (taken, of course, under current conditions). This RP data may be used with the other data in the estimation of the model parameters. Utility for the d-th actual Green Bay fishing day is given by:

$$U_{id}^{G} = b' x_i^{G} + e_{id}^{G} \tag{8}$$

where x_i^G is a vector of characteristics of Green Bay under actual conditions, and e_{id}^G is a random component with variance s_G^2.

In deciding how many days to fish Green Bay, the individual compares utility at Green Bay to utility at other sites, given by Equation 4. For RP data we assume the individual knows his random component at the time each fishing day's choice is made, so that the probability of going to Green Bay on day d is:

$$\begin{aligned} P_i^G &= P(U_{id}^G > U_{ij}^0) \\ &= P(b' x_i^G + e_{id}^G > b' x_{ij}^0 + e_{ij}^0) \\ &= P(e_{ij}^0 - e_{id}^G < b' x_i^G - b' x_{ij}^0) \\ &= \Phi[(b' x_i^G - b' x_{ij}^0) / s_{0-G}] \end{aligned} \tag{9}$$

where:

$$s_{0-G}^2 = \mathrm{Var}\,(e_{ij}^0 - e_{ij}^G) = s_0^2 + s_G^2 - 2 s_{G0} \tag{10}$$

7. Note that in Equation 6, b appears twice. On one occasion it is normalized by $\sqrt{2} s_e$, and on the other by s_{0-e}. Also note that under the alternative assumption that disturbances are known to the individual a priori, he would perform the conceptual experiment of generating n_i pairs of disturbances, evaluating utility under the two scenarios, and counting the number of Green Bay trips under the assumption of utility maximization. This process would also imply Equations 5-7.

Since P_i^G is a function of b, the information contained in n_i^G is useful in estimation, and is incorporated into the likelihood.

To summarize, the random variable e takes a variety of identifying notations. Table A-1 summarizes the cases.

Table A-1
Summary of Disturbances

Disturbance	Site	Data Type
$e_{ij}^{k_{ij}}$	Green Bay, proposed	SP
e_{ij}^{0}	Other than Green Bay	SP/RP
e_{ij}^{G}	Green Bay, actual	RP

Disturbances for revealed versus stated preference data may or may not have different variances, which would mean the informational content (toward the estimation of b) would differ. Since we allow correlation among disturbances and can only estimate the variance of the disturbance differences, we cannot assess the relative information content of the different kinds of data.

A.5 THE LIKELIHOOD FUNCTION

The maximum likelihood parameter estimates are consistent. They are also asymptotically efficient under the additional assumption that e_{ij}^{0} and $e_{ij}^{k_{ij}}$ are uncorrelated across *j*. The likelihood function is a function of the probabilities of the preferred alternatives from the Green Bay pairs (Section A.2), the conditional probabilities for how often the preferred Green Bay alternatives would be selected versus a non-Green Bay site using RP data on the number of total fishing days (Section A.3), and the probabilities for actually visiting Green Bay using RP data on the number of Green Bay days (Section A.4). The likelihood function is:

$$L(n_{ij}, k_{ij}, n_i^G, i = 1,\ldots,m, j = 1,\ldots,J \,|\, x_{ij}^1, x_{ij}^2, n_i; b, s_0, s_e, s_G) =$$
$$\prod_{i=1}^{m}\left[\binom{n_i}{n_i^G}(P_i^G)^{n_i^G}(1-P_i^G)^{n_i-n_i^G}\prod_{j=1}^{J}P(N_{ij}=n_{ij}\,|\,K_{ij}=k_{ij})P(K_{ij}=k_{ij})\right] \quad (11)$$

Note that in this likelihood b appears in several expressions: in P_i^G normalized by s_{0-G}, in $P(N_{ij} = n_{ij} | K_{ij} = k_{ij})$ normalized by s_{0-e}, and in P_i and $P(K_{ij} = k_{ij})$ normalized by $\sqrt{2}s_e$. The ratios of any two of these three parameters are identified in estimation.[8]

A.6 DERIVATION OF EQUATION 6

Consider the probability of choosing Green Bay site k_{ij} over a non-Green Bay site, conditional on the choice of Green Bay site k_{ij} over Green Bay site 3 - k_{ij}. To ease the notation, suppose alternative 1 is chosen rather than alternative 2, and the individual and choice occasion subscripts are ignored. Under assumptions 2 and 3, the random vector (e^1, e^2, e^0) has a multinormal distribution with zero mean vector and covariance matrix:

$$\begin{pmatrix} s_e^2 & 0 & s_{eo} \\ 0 & s_e^2 & s_{eo} \\ s_{eo} & s_{eo} & s_0^2 \end{pmatrix} \tag{12}$$

This implies:

$$w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} e^0 - e^1 \\ e^2 - e^1 \end{pmatrix} \sim N(0, \Omega) \tag{13}$$

where:

$$\Omega = \begin{pmatrix} s_{0-e}^2 & s_e^2 \\ s_e^2 & 2s_e^2 \end{pmatrix} \tag{14}$$

The probability in Equation 6 is a conditional probability of a bivariate normal random variable, where the conditioning event does not have zero probability (which is the more usual case).[9] Let $a_1 = -b'(x_{ij}^0 - x_{ij}^{k_{ij}})$ and $a_2 = -b'(x_{ij}^{3-k_{ij}} - x_{ij}^{k_{ij}})$. From Amemiya (1994, pp. 35-36), denoting the joint, marginal, and conditional density functions of w and its elements as f, we have:

---

8. Although all parameters are listed separately, it is evident that normalizations are necessary.

9. It *is* a conditional probability, rather than a conditional expectation, so the Mill's ratio results from the selection literature (e.g., Maddala, 1983, p. 367) cannot be used.

$$f(w_1|w_2 < a_2) = \frac{\int_{-\infty}^{a_2} f(w_1,w_2)dw_2}{P(w_2 < a_2)} \tag{15}$$

so that:

$$P(w_1 < a_1|w_2 < a_2) = \int_{-\infty}^{a_1} f(w_1|w_2 < a_2)dw_1 = \frac{\int_{-\infty}^{a_1}\int_{-\infty}^{a_2} f(w_1,w_2)dw_2dw_1}{\int_{-\infty}^{\infty}\int_{-\infty}^{a_2} f(w_1,w_2)dw_2dw_1} \tag{16}$$

This is the ratio of a bivariate normal cumulative distribution function evaluated at a_1 and a_2 to a univariate normal cumulative distribution function evaluated at a_2:

$$P(w_1 < a_1|w_2 < a_2) = \frac{\Phi_2\left(a_1 / s_{0-e}, a_2 / \sqrt{2}s_e; r\right)}{\Phi\left(a_2 / \sqrt{2}s_e\right)}, \tag{17}$$

which is Equation 6.